AD-A252 295

REPORT DOCUMENTATION PAGE

Form Approved
OPM No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 1988	3. REPORT TYPE AND DATES COVERED Unknown

4. TITLE AND SUBTITLE Uncertainty and the Conditioning of Beliefs'	5. FUNDING NUMBERS DAAB10-86-C-0567
6. AUTHOR(S) Henry E. Kyburg, Jr.	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Rochester Department of Philosophy Rochester, NY 14627	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army CECOM Signals Warfare Directorate Vint Hill Farms Station Warrenton, VA 22186-5100	10. SPONSORING/MONITORING AGENCY REPORT NUMBER 92-TRF-0003

DTIC ELECTE MAY 13 1992 S D D

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT Statement A; Approved for public release; distribution unlimited.	12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)* Uncertainty is part of the human condition. Whether we will or no, we must act, we must amke decisions, in the face of uncertainty. Some authors have proposed that uncertainty be regarded as essentially a subjective matter. Our first goal is to draw the teeth of the classical subjectivistic argument that one must be prepared to meet all bets on the basis of one's "degrees of belief." The Dutch book theorem, which purports to have this as a consequence, is stated and criticized. Other criticisms of logical and subjective probability are considered. This leads to the consideration of alternative conceptions of how to represent epistemic uncertainty. A variety of alternative have been offered, including, recently, Glenn Shafer's theory of belief functions. An exposition of Shafer's theory is offered. We then relate Shafer's theory of belief functions to a theory that represents (and updates) uncertainty in terms of convex sets of classical probability functions. Finally, we discuss the question of the decision principles that can be employed in the cas of both the convex set representation and the belief function representation of uncertainty.

14. SUBJECT TERMS Artificial Intelligence, Data Fusion, Uncertainty of Beliefs, Conditioning of Beliefs	15. NUMBER OF PAGES 18
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFED	UNCLASSIFIED	UL

CHAPTER 4

UNCERTAINTY AND THE CONDITIONING OF BELIEFS[1]

Henry E. Kyburg, Jr.

Abstract--Uncertainty is part of the human condition. Whether we will or no, we must act, we must make decisions, in the face of uncertainty. Some authors have proposed that uncertainty be regarded as essentially a subjective matter. Our first goal is to draw the teeth of tne classical subjectivistic argument that one must be prepared to meet all bets on the basis of one's "degrees of belief." The Dutch book theorem, which purports to have this as a consequence, is stated and criticized. Other criticisms of logical and subjective probability are considered. This leads to the consideration of alternative conceptions of how to represent epistemic uncertainty. A variety of alternatives have been offered, including, recently, Glenn Shafer's theory of belief functions. An exposition of Shafer's theory is offered. We then relate Shafer's theory of belief functions to a theory that represents (and updates) uncertainty in terms of convex sets of classical probability functions. Finally, we discuss the question of the decision principles that can be employed in the case of both the convex set representation and the belief function representation of uncertainty.

- - - - -

I. BACKGROUND

It is a fact of life, whether we applaud it or deplore it, that we must decide and act in the face of uncertainty and on the basis of

[1]Research on which this work is based was partially supported by the U.S. Army Signals Warfare Center.

incomplete information. It is argued by some philosophers that if we had complete information, we would not have to act in the face of uncertainty; but it has also been argued by others (and by philosophers of quantum mechanics in particular) that even if we had full information, we would not be able to eliminate uncertainty.

It is true that if complete knowledge included knowledge of the future we would not have to face uncertainty if we had complete knowledge. We mortals have been seeking that kind of knowledge for centuries: in the stars, in chicken entrails, in science. We have not found it. When, long ago, we believed that the gods knew the future and told the truth, we also understood that the oracles spoke in riddles. Thus, though we had been told what the future would bring, our interpretation of what we had been told introduced a new level of uncertainty.

Although probability theory has developed only in very recent historical times, people have had some understanding of the practical aspects of uncertainty for as long as gambling has been a pastime. It is an important and interesting modern question to ask to what extent the gambling model of action in the face of uncertainty has general validity.

A number of authors, philosophers, statisticians, and probability theorists, have drawn a distinction between the kind of uncertainty that characterizes our general knowledge of the world, and the kind of uncertainty that we discover in gambling. This distinction appeals to intuition. If the dice are fair, the chances of two ones on a roll of two dice is 1/36. But what is the chance that the dice are fair? That seems quite a different question.

Taking the distinction seriously has led to two dominating views concerning probability and uncertainty. One identifies probability with long-run relative frequency. This view was given explicit articulation by John Venn (1866)--though Aristotle, who said that what was probable was that which happened for the most part, might also be taken as a frequency theorist. Its best known advocate was the positivist, Richard von Mises (1928). This view appears to account for the assessment of the chances of getting the sum two in a roll of two dice: that result happens about 1/36 of the time in the long run; therefore the probability should be taken to be 1/36.

On the other hand, dice can be more or less fair; and there is no well established and agreed upon relative frequency with which dice are

unfair. This seems to be quite a different problem. And so a different conception of probability has been devised to deal with it.

More accurately, *two* different conceptions. For an early hope, articulated by John Maynard Keynes (1921), was that it should be possible to define a logical conception of probability that would measure the degree of uncertainty of any hypothesis on any evidence. (Others had already conceived of a notion of probability that would be epistemic--i.e., that would determine the *rational* degree of belief of an agent possessed of given evidence.) Thus, given what we know about dice, social customs, physics, the interests of our friends, the state of the economy, and so on, there would be *one* logically fixed probability for the hypothesis that a given die is loaded to a given degree. Keynes supposed that there was such a probability, fixed and determined by background knowledge and evidence, but he did not assume that it was a real number in the interval [0, 1]. In particular, he thought there was good reason to suppose that sometimes probabilities were not comparable: taking what I know about the world as evidence, I cannot say whether rain tomorrow is more probable than, less probable than, or as probable as, the occurrence of heads on the next toss of this coin. This is not through any failure of logical insight, or weakness of intellect. It is simply that the abstract objects we call "probabilities" are not simply ordered, but only partially ordered. They form a lattice, whose supremum and infimum are 1 and 0, but in which there are many non-comparable pairs.

The idea of a lattice of probability values was pursued briefly by B. O. Koopman (1940), and then disappeared until the late 1950's, to be revived under a different name.

Meanwhile, a number of writers continued to pursue the idea of probability as a logical relation. Foremost among these was Rudolf Carnap (1950). The idea here is this: given a formal language, there is an intuitively correct assignment of real-valued measures m to its sentences such that if h is an hypothesis, and e is our total store of evidence, the probability--legislative for rational belief--of h conditional on e is $m(h \ \& \ e)/m(e)$. This is Keynes' vision, formalized and simplified by the assumption that probabilities are real numbers in the unit internal.

Others (e.g. Harold Jeffreys, 1939; Jaakko Hintikka, 1966; Ilkka Niiniluoto, 1976) have pursued this vision. It has turned out to be a complicated job to assign measures to all the sentences of a complicated

and general language. The only feasible way of doing it seems to be to parametrize the language (number of one-place predicates, impact of canonical evidence on a canonical assertion, etc.). But then, if the assignment is to be "rational", and defensible as rational, we must ask the question: Why should these parameters have the values we have given them? The answers have been hard to find.

Shortly after Keynes had proposed his logical view of probability, Frank Ramsey, a colleague of Keynes' at Cambridge, criticized it from the point of view of what has come to be called personalistic Bayesianism (Ramsey, [1931] 1950). Ramsey argued that there was no point in saying that something was "legislative for rational belief" unless you could measure belief. Ramsey devised a pragmatic (or operational) method for measuring belief according to which there was a clear argument (we leave aside here the question of its validity) that beliefs should be real valued and should conform to the probability calculus. He could find no argument that they should satisfy any other constraints. Thus he rejected the logical conception of probability in favor of a subjectivistic conception.

Bruno de Finetti (1937), and L. J. Savage (1954), both statisticians, also endorsed the view that such probabilities as the probability that the die is biased, could only be subjective. Of course this is not to say that such probabilities do not depend on evidence; it is only to say that it is some individual who evaluates the evidence, and that there is no reason that you and I should both evaluate the same evidence in the same way. From their views a very lively tradition has evolved. It is called "Bayesianism", though it is not Bayes theorem that is at issue.

Bayes theorem is a theorem of the conventional probability calculus. It says that the probability of a hypothesis, relative to some evidence, is the prior probability of that hypothesis, multiplied by the probability of the evidence on the supposition that the hypothesis is true, divided by the prior probability of the evidence. If we can suppose that we have a number of exhaustive and exclusive alternatives that can be taken as hypotheses, the prior probability of the evidence can be taken as a sum of terms consisting of the prior probability of an alternative hypothesis, multiplied by the conditional probability of the evidence, given that hypothesis. Since it is generally (but perhaps ill-advisedly) supposed that the probability of a piece of evidence, given a statistical hypothesis concerning evidence of that sort, is unproblematic,

the serious question for the Bayesian point of view is the source and status of the prior probabilities of the hypotheses.

Ramsey's solution is that rationality imposes no constraint. A man may have whatever degrees of belief he will, provided only that they satisfy the constraints of the probability calculus. Some writers suppose that prior probabilities are determined by some general principle (e.g., the maximum entropy, or least information, principle --E. T. Jaynes, 1968), but the application of the general principle depends on the "formulation of the problem," which is again a relatively subjective matter. Logical theorists, as already noted, require the specification of parameters in order to determine the prior probabilities of hypotheses.

By Ramsey's Dutch book argument, these are all the alternatives there are. Ramsey's argument is that you should have degrees of belief such that you could accept all bets offered at odds corresponding to your degrees of belief without having a Dutch book--a set of bets that entails that you lose whatever happens--made against you.

It follows that probabilities are real-valued. And it follows that they must be updated by Bayes theorem: i.e. that there must be prior probabilities for every hypothesis. But these probabilities must then be subjective (Ramsey's view) or they must be obtained systematically, according to general principles (the logical view, the maximum entropy view). But in the latter cases there are important parameters that are just as subjective as Ramsey's degrees of belief.

To avoid this conclusion, and the arbitrariness it embodies, we must draw the teeth of Ramsey's argument (or find compelling rational principles that do not require subjective judgment).

Although the issues involved can be complex (see Fahiem Bacchus, Kyburg, and Mariam Thalos, 1989), the basic idea is simple. To be sure, it is irrational to accept a set of bets according to which you lose something you value no matter what happens. But this fact about rationality says nothing about degrees of belief. The crucial connection to degrees of belief is the part of the argument that identifies one's degree of belief in a statement S with the least odds at which one would bet on S. But it is not at all obvious that one has degrees of belief, or that they are associated with the odds at which one is willing to bet in the way that Ramsey suggests.

Specifically, while it seems reasonable to say that the least odds at which I am willing to bet on S represent a kind of lower bound of my

belief in S, and similarly for the greatest odds at which I am willing to take a bet on S, it is not at all obvious that these two sets of odds should be complementary. If I am unsure about S, I may well offer odds of 1 to 2 on S, and odds of 1 to 2 against S, without being willing to offer any intermediate odds on either.

Another approach to determining the basic properties of probability is the analytic approach exemplified by Richard T. Cox (1961). It turns out that the most innocuous and harmless-sounding conditions imposed on uncertainty can be shown to lead directly to the conventional probability calculus. Among these conditions is, of course, something akin to simple order among probabilities.

One should remember, at this point, the basic distinction that has given rise to these problems: the distinction between probabilities that can be construed as frequencies, and probabilities that cannot be so construed. We shall see later (in section IV) that this is not as simple a distinction as it appears to be.

II. VARIANTS ON PROBABILITY

There are a number of objections to the classical probability calculus as a representation of uncertainty. Among them are these:

> 1. Strictly speaking, frequencies only apply to classes or predicates. One can speak of the frequency of heads on tosses of this coin, but not usefully of the frequency of heads on the next toss of this coin.
> 2. Many of the events whose probability we wish to speak of (the probability that an individual exhibiting a unique background and cluster of symptoms has a certain disease) are not related in any obvious way to statistical knowledge.
> 3. Subjective and logical interpretations of probability give us numbers, but they are arbitrary. The numbers provided by a logical view reflect arbitrary general assignments to the sentences of an artificial language.

> The numbers provided by a subjectivistic view may (for all the theory can say) reflect mere whimsy.
>
> 4. None of the theories provides a representation that can indicate directly that a probability is unknown or poorly known: that is, that can indicate the difference between the probability of heads on the next toss of a well tested coin, and the probability of heads on a totally unknown coin: both may be represented by the number 0.500. (The difference is indicated indirectly by the conditional probability of heads given heads).
>
> 5. Bayesian and logical views often require the assignment of probabilities to a great many entities. Thus in computing the conditional probability of H given E, we may require the probability of E on every alternative hypothesis to H.

A number of philosophers, including Karl Popper (1959), Nicholas Rescher (1958), Carl Hempel and Paul Oppenheim (1945), have offered measures of evidential or factual support. These measures are not probabilities, though they are relatively simple functions of probabilities. (For a table exhibiting their relations, and the ways in which they are related to conventional probability measures, see Kyburg, 1970.)

These measures are designed explicitly to guide our beliefs with respect to general hypotheses: e.g., the hypothesis that the die is biased in a certain way, the hypothesis that all A's are B's, the Newtonian hypothesis (or theory) governing celestial motions, the hypothesis that less than 30% of the A's are B's. Of course these are exactly the sorts of hypotheses whose probabilities one needs to feed into Bayes theorem. What happens when we use these numbers in a decision theoretic context?

As soon as we try to use such measures in a decision theoretic context, Ramsey's (or Cox's) arguments apply full force. Here we have no question of merely representing the open and vague and ambiguous notion of belief; here we have a straightforward matter of decision involving (presumably) well specified utilities. It may well be that my psychological state concerning whether drug *A* will relieve the symptoms of patient *P* is best represented by a vector. But that is another matter.

The most intuitive way of associating degrees of belief with numbers, employed by Savage (1954), is this: What is the most you would pay for a ticket that would yield a dollar if S is true? That is your probability for S. On all of these variant views, the support of a hypothesis is supposed to be real-valued, and normalized to the [0,1] interval. If the numbers can be used to weight utilities, in a decision theoretic context, then it follows from Ramsey's arguments (among others) that they must satisfy the axioms of the probability calculus. That is, the measures purporting to be variants on probability cannot be viable if they lead to a book being made against one. Or they cannot be taken as guiding our decisions in the face of uncertainty.

A similar story may be told about Artificial Intelligence. Expert systems, it is clear, must be capable of handling uncertainty. Various systems have employed various representations of uncertainty. For example, MYCIN (E.H. Shortliffe, 1976) is an expert system designed to provide assistance in medical diagnosis. The certainty factors of MYCIN, for example, range from -1.0 to 1.0, where -1.0 applied to S represents full confidence that S is false, and 1.0 applied to S means full confidence that S is true. In the process of inference, certainty factors are combined according to special rules.

Certainty factors are not probabilities. Not only is the range wrong, but the rules of combination are inconsistent with the (Bayesian) rules for the combination of probabilities. If they were to be used as weighting factors in making decisions, in the same way that probabilities are used, Ramsey's arguments could be used to show that the decisions would not be rational: in a sense, the physician could have a book made against him. (There is no suggestion that certainty factors *should* be used this way; there is no suggestion of computing expectations based on certainty factors and using these expectations for arriving at decisions. But the Dutch book argument provides a *reason* for eschewing these suggestions.)

Another approach to the treatment of uncertainty that has received much attention in artificial intelligence is Shafer's (1976) theory of belief functions. This is a clear mathematical theory, based on earlier work of Arthur Dempster (1967; 1968). It is designed to overcome some of the discomforts that people have felt concerning both the subjectivistic Bayesian theories and their logical variants, as well as frequency theories.

III. BELIEF FUNCTIONS

The basic building block of the theory of belief functions is the *frame of discernment* Ω. A frame of discernment may be thought of as a set of possible worlds, to use philospher's jargon, but they need be construed in no more detail than concerns us in a given context. If I am concerned with the outcome of a coin toss, there are only two possible worlds that concern me: for example, one in which the coin lands heads, and one in which it lands tails.

My beliefs are represented by an assignment of mass to *sets* of possible worlds, including the possibility of assigning mass to unit sets or singletons of possible worlds, and the possibility of assigning mass to the set of all possible worlds. Masses are non-negative real numbers between 0 and 1. The total mass assigned is 1.0.

A belief function, or support function, is a function whose domain is sets of possible worlds (subsets of Ω), and whose values lie in $[0,1]$. For $X \subset \Omega$, $\mathrm{Bel}(X) = \Sigma m(A)$, where $m(A)$ is the mass assigned to $A \subset \Omega$, and the summation extends over all subsets A of X, including X itself.

$\mathrm{Bel}(X)$ represents the amount of belief I have in the possibility X. It is one of the attractive features of this system, as opposed to classical probability systems, that I can have very little belief in X and at the same time very little belief in its denial, which we denote by '$\sim X$': that is, instead of $P(\sim X) = 1 - P(X)$, we can have both $\mathrm{Bel}(X) = \epsilon$ and $\mathrm{Bel}(\sim X) = \epsilon$. To express complete ignorance about everything, we can assign a mass of 1.0 to Ω, and a mass of 0 to every proper subset of Ω.

It is easy to see how attractive this can be. Somehow, to know the probability of something is to know *something*; a probability of 0 represents, not ignorance, but certainty just as much as a probability of 1. But a probability of a half doesn't seem to represent ignorance, either. In the new system, belief in S equal to 0 *may* represent ignorance; it does so if belief in the denial of S is also 0.

Let us now consider updating--the way belief functions and mass functions change with the accumulation of evidence. "Evidence" is construed as a frame of discernment with a belief function defined on it. This represents what has happened to us--what we are taking account of. If Ω contains six subsets corresponding to the outcome of a toss of

a slightly suspicious die, it might have masses of 0.1 on each of those subsets and a mass of 0.4 (representing ignorance) on Ω itself. Now let us suppose we are told by a person of doubtful reliability that the toss resulted in an odd number of spots. This might be represented as the same frame of discernment with a mass of 0.7 on the set corresponding to odd tosses, and 0.3 on Ω.

Our beliefs should now be represented as the result of *combining* these two belief functions. (Note that we have not required that the "evidence" be known with certainty.) The procedure is to consider all the subsets of Ω that have mass according to either belief function; if S bears positive mass $m_1(S)$ according to the first belief function, and T bears positive mass $m_2(T)$ according to the second belief function, then we assign a mass of $m_1(S) \times m_2(T)$ to the intersection of S and T, provided that intersection is not empty. If it is empty--i.e. if it represents an impossible state of affairs, such as the toss landing two and also being odd--then we assign it 0.0. To account for this lost mass and to get back to a canonical belief function, we normalize by dividing each number by 1-k, where k is the sum of the products of the mass of subsets that are inconsistent with each other.

Thus we have, for our example: the mass assigned to the intersection of 'one' and 'odd' is 0.1x0.7 = 0.07; the mass assigned to the intersection of 'one' and Ω is 0.1x0.3 = 0.03; etc., all normalized to take account of the impossibility of certain intersections. The following table illustrates the procedure.

	odd	Ω
1	0.1x0.7	0.1x0.3
2	0.0	0.1x0.3
3	0.1x0.7	0.1x0.3
4	0.0	0.1x0.3
5	0.1x0.7	0.1x0.3
6	0.0	0.1x0.3
Ω	0.4x0.7	0.4x0.3

The normalizing number is 1 - 3x0.07 = 1 - 0.21 = 0.79. Thus we find that the belief we should attribute to 'three' is (0.1x0.7 + 0.1x0.3)/0.79

= 0.127; the belief we should attribute to odd is 0.734;[2] the belief we should attribute to Ω (ignorance) is 0.152; etc.

There is a special case that corresponds to Bayesian conditionalization. If our evidential belief function assigns mass 1.0 to a single subset of Ω (and perforce 0 to every other subset of Ω), then we may compute the *updated* probability of any subset of Ω by means of what Shafer (1976, p. 67) calls "Dempster's rule of conditioning." In this X is arbitrary, and B is the set corresponding to the evidence (we assume that the belief function for "not B" is positive, $Bel(\sim B) > 0$):

$$(1)\quad Bel(X \mid B) = [Bel(X \cup \sim B) - Bel(\sim B)]/[1-Bel(\sim B)].$$

A *simple support function* is a belief function that results from the assignment of mass to Ω and to a single subset of Ω. A *separable support function* is a belief function that results from the combination of a finite number of simple support functions. There are other support functions, and indeed there are belief functions that are not support functions, but the separable support functions represent quite a broad class. It is therefore of interest to note that there is a procedure for expanding Ω so that the result of updating by a *simple* support function can be represented as an instance of Dempster's rule of conditioning (Kyburg, 1987). It follows that updating by a separable support function can be represented by a sequence of steps of Dempster conditioning.

We have a general and attractive procedure for representing and updating uncertainty here. It seems quite different from probability. But one of the differences is not so nice: there is no obvious decision procedure based on belief functions. In the case of any standard subjective or logical probabilistic approach, we can apply the principle of maximizing expected utility to decision theory. Here we cannot.

---

[2] The measures assigned to 1, 3, and 5 are each 0.1x0.7 + 0.1x0.3, or a total of 0.30, plus the measure assigned to the general class, odd, by the new information, multiplied by the non-specific assignment provided by the old information, 0.7x0.4. This sum, 0.58, is normalized by dividing by 0.79, which yields 0.734.

IV. BELIEF FUNCTIONS AND PROBABILITIES

Dempster (1967; 1968) originally referred to "upper" and "lower" probabilities. The idea, but not the terminology is preserved in Shafer: the belief function gives the lower probability; there is a dual notion, plausibility, that corresponds to an upper probability. (The plausibility of X, Pl(X), is defined to be 1 - Bel(~ X).)

First, note that the space Ω of possibilities is just another way of representing propositions or statements. A subset of Ω corresponds to a statement. A probability function defined over Ω consists in the assignment of a number to each complete description of a state of affairs--or each atomic possible world. A set of possible worlds, corresponding to a disjunction of the atomic world descriptions, will then receive as its measure the sum of the numbers assigned to its atoms. The translation between statements and subsets of Ω is straightforward.

Shafer's system does not *require (but it allows)* the assignment of masses to the singletons (corresponding to the atomic worlds). We can capture this aspect of the system by considering, not a single assignment to the atomic worlds, but a set of assignments. For example, consider a simple frame of discernment containing two states of affairs: heads and tails. The subsets consist of ∅, which has mass 0, H = {heads}, T = {tails}, and Ω = {heads, tails}. Let us, to reflect our uncertainty about the coin, assign mass 0.4 to H and to T, and mass 0.2 to Ω. We can accomplish the same thing with a set of probability functions: we can consider the set of all those classical probability functions whose domain is {heads, tails}, and whose value for heads lies between 0.4 and 0.6. For every function P in this set, P(Tails) = 1 - P(Heads). Belief and plausibility are now most naturally thought of as lower and upper probabilities, respectively.

This holds quite generally. Given any belief function defined on a frame of discernment, there will exist a set of classical probability functions, defined on the same set of possible worlds, with the property that for any subset X of the frame of discernment, the belief assigned to X, Bel(X), is the minimum of the values P(X) for probability functions P in that set, and the plausibility assigned to X, Pl(X) = 1 - Bel(~ X), is the maximum. Furthermore, the set of probability functions with this property is convex: If P and Q belong to the set of probability functions

in question, so does the function PQ, where $PQ(X) = a(P(X)) + (1-a)(Q(X)), 0 \leq a \leq 1$.

Surprisingly, the converse relation does *not* hold. There are sets of probability functions to which there corresponds no belief function. Furthermore, these examples need not be bizarre.

Consider a compound experiment[3] consisting of performing a mixture, in unknown ratio *p*, of two experiments: (1) tossing a fair coin twice, or (2) drawing a coin from a bag containing 60% two-headed and 40% two-tailed coins, and tossing it twice. The outcomes of the compound experiment that interest us are A, the event that the first toss lands heads, and B, the event that the second toss lands tails. CP is to be the convex set of possible distributions of outcomes on the compound experiment. $CP = \{< \frac{1}{4}p + 0.6(1-p), \frac{1}{4}p, \frac{1}{4}p, \frac{1}{4}p + 0.4(1-p) > p \in [0,1]\}$. This is a set of quadruples. The first parameter is the frequency of HH, the second of HT, the third of TH, and the fourth of TT, on an arbitrarily large number of repetitions of the compound experiment. We are representing our knowledge of the long-run outcomes of the experiment by a convex set of probability distributions. We call this the *convex set representation.*

Let $P_m(X)$ be the least value of $P(X)$ for $P \in CP$. Then $P_m(A \cup B) < P_m(A) + P_m(B) - P_m(A \cap B)$. We would like to identify $P_m(X)$ with $Bel(X)$. But one of Shafer's (1976, pp. 38-39) theorems requires that for all belief functions, $Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B)$. This shows that P_m *cannot* be a belief function. We cannot represent this uncertain situation by belief functions, but the convex set representation is quite straightforward and intuitive.

Both representations are of interest, however. The belief function representation is an easy one to manipulate; the convex set representation is difficult to deal with computationally. The convex set representation is intuitively clear; the belief function representation seems artificial. Furthermore, the two representations are mutually enlightening.

As an example, let us consider updating in the light of new evidence. In the convex set representation, we can represent classical Bayesian conditionalization. Given a single probability function P, the conditional probability of a hypothesis H on evidence E, when P(E) >

[3]This example was suggested by Teddy Seidenfeld in conversation.

0, is $P(H \mid E) = P(H\&E)/P(E)$. If E represents our total increment of evidence, the principle of *confirmational conditionalization* (Isaac Levi, 1980) directs us to adopt as our credence function, $P(H) = P(H\&E)/P(E)$.

Given a set of probability distributions CP, we can accomplish the same end. Let CP be a convex set of classical probability functions. Let our total new evidence be E. Then our new belief state should be represented by CP′, where CP′ is the set of probability functions of the form $P(H\&E)/P(E) = P(H \mid E)$, for P in the set CP, and $P(E) > 0$. It turns out that when CP is convex, and there is at least one function P in CP such that $P(E) > 0$, CP′ is convex, too.

Now a belief function can be represented by a convex set of probability functions (but not vice versa), and, when E is a piece of evidence we learn for certain, we can apply both Dempster conditioning and confirmational conditionalization. It turns out that Dempster conditioning imposes tighter constraints on our degrees of belief than does confirmational conditionalization. Writing $Bel(X \mid Y)$ for the updated belief function and $Pl(X \mid Y)$ for the updated plausibility function, we have the following relation, where the infimum (inf) and supremum (sup) are taken over the set of functions CP′ (Kyburg, 1987):

(2) $\inf P(H \mid E) \leq Bel(H \mid E) \leq Pl(H \mid E) \leq \sup P(H \mid E)$,

In this, equality holds only in rather special cases, when certain distributions are ruled out as impossible by *all* the P's in CP.

One response to this fact would be to be pleased that the belief function form of updating leads to "stronger" results than generalized Bayes. I believe that this response would be mistaken. We have given no specific interpretation to the members of CP. In particular, they may be purely objective chances or frequencies, or they may (as I would usually construe them) be epistemic probabilities directly based on knowledge of frequencies or chances. In either case, inf P(X) *can* represent the value of a frequency or a chance. In adopting $Bel(H \mid E)$ as your odds-determining measure, you may be ruling out this possibility groundlessly. This corresponds to a well known difficulty in the theory of belief functions--namely, that very ambiguous evidence can lead to

very unambiguous belief functions, in which Bel(X) = Pl(X) (see Lotfi Zadeh, 1979).

I refer to this as a difficulty, but of course whether it is or not depends in part on what is at stake. One can imagine circumstances in which the greater precision afforded by Dempster conditioning more than offsets the security provided by conditionalization. For example, in a situation in which the agent is forced to make book with all comers, and in which the real distributions in the world are unimodal, *and* in which the decision rule has any of a number of plausible forms, it is clear that someone following Dempster conditioning will probably (!) come out ahead of someone who follows classical conditioning.

We can also raise the question of whether or not conditionalization is itself rational. There have been a number of arguments in favor of confirmational conditionalization (Paul Teller, 1976; Bas van Fraassen, 1984). We do not find these arguments persuasive, and in fact have argued against them in Kyburg (1987) and in Bacchus et al. (1989).

But what are the plausible forms of a decision rule? The relation between a representation by convex sets of probability functions and Shafer's representation by belief functions gives us a handle on this question, but it is by no means settled.

V. *DECISION THEORY*

One of the most attractive features of classical probability--and indeed what the whole approach of subjectivistic probability is based on --is that it lends itself to a very simple and persuasive decision rule: *Maximize Thy Expected Utility.* At the same time, one of the interesting aspects of any alternative to a single classical probability function as a representation of belief is the way in which it lends itself to some form of decision theory.

The close relation between belief functions and convex sets of classical probability functions suggest relations between the decision theory appropriate for sets of classical probability functions and the decision theory appropriate for belief functions. But what is the decision theory appropriate for sets of probability functions?

In the first place we can apply the classical Bayesian procedure. When we have intervals of probability, we can consider the maximum expected utility and the minimum expected utility of one decision, and the maximum and minimum expected utility of another decision. If the minimum expected utility of one decision exceeds the maximum expected utility of another decision, we have a clear ordering of those two decisions.

More generally and more precisely, let us say one decision *dominates* another when the *minimum* expected utility of the first exceeds the *maximum* expected utility of the second. In that case we clearly have nothing to lose if we forget about the second possibility. So, on perfectly classical grounds, we can ignore dominated alternatives.

Beyond this, the decision theory for convex sets of classical probability functions reflects classical decision theoretical problems. It is a theory that should take account of indeterminacy (as opposed to uncertainty), but how to do this is an open question. In classical terms, if X and Y are outcomes that are both possible and the utility of action A exceeds that of action B if X is the case, but the opposite holds if Y is the case we are faced with an indeterminate situation, unless we know the probabilities of the alternatives that X is the case and that Y is the case.

In such cases there are various rules that one might apply. Minimax is one, minimax regret another. Levi (1980) has explored a lexical approach based on a sequence of notions of admissibility. There are no doubt any number of alternatives, almost none of which have been adequately discussed. It is not my purpose here to defend one particular approach to decision under these circumstances, but merely to point out the relevance of classical decision theory to the case in which uncertainty is represented by belief functions. The claim that there is no decision theory to go with the uncertainty representation of belief functions is clearly wrong. But there is no decision theory for these cases on which all reasonable persons agree.

No more, of course, have the classical issues of decision in the face of uncertainty been solved. But it is significant that for the classical problem there are a number of alternatives that are considered worthy of serious discussion. Equally, for the convex probability case, or the belief function case, these alternatives should receive serious consideration. It is hoped that further consideration will reveal some principles that will enlighten our decision-theoretic concerns. In any

ent it is clear that there is a decision-theoretic framework that is pplicable in the belief function framework, and it is also clear that its pplication is not a matter whose principles are entirely settled.

REFERENCES

Bacchus, Fahiem; Kyburg, Henry and Thalos, Mariam (1989), "Against Conditionalization," TR256, Computer Science, University of Rochester.

Carnap, Rudolf (1950), *The Logical Foundations of Probability*, Chicago: University of Chicago Press.

Cox, Richard T. (1961), *The Algebra of Probable Inference*, Baltimore: Johns Hopkins Press.

de Finetti, Bruno (1937), "La Prevision: Ses Lois Logiques, Ses Sources Sujectives," *Annales De L'Institute Henri Poincare*, 7, pp. 1-68.

Dempster, Arthur P. (1968), "Upper and Lower Probabilities Generated By a Random Closed Interval," *Annals of Mathematical Statistics*, 39, pp. 957-66.

Dempster, Arthur P. (1967), "Upper and Lower Probabilities Induced By a Multivalued Mapping," *Annals of Mathematical Statistics*, 38, pp. 325-39.

Hempel, Carl, and Oppenheim, Paul (1945), "A Definition of 'Degree of Confirmation'," *Philosophy of Science*, 12, pp. 98-115.

Hintikka, Jaakko (1966), "A Two-Dimensional Continuum of Inductive Methods," in Jaakko Hintikka and Patrick Suppes, eds., *Aspects of Inductive Logic*, Amsterdam: North Holland, pp. 113-32.

Jaynes, Edward T. (1968), "Prior Probabilities," *IEEE Transactions on Systems Science and Cybernetics*, 4, pp. 227-41.

Jeffreys, Harold (1939), *Theory of Probability*, Oxford: Oxford University Press.

Keynes, John Maynard (1921), *A Treatise on Probability*, London: Macmillan.
Koopman, Bernard O. (1940), "The Axioms and Algebra of Intuitive Probability," *Annals of Mathematics*, 41, pp. 269-92.
Kyburg, Henry E., Jr. (1987), "Bayesian and Non-Bayesian Evidential Updating," *Artificial Intelligence*, 31, pp. 271-93.
Kyburg, Henry E., Jr. (1970), *Probability and Inductive Logic*, New York: Macmillan.
Levi, Isaac (1980), *The Enterprise of Knowledge*, Cambridge: MIT Press.
Mises, Richard von (1928), *Probability, Statistics and Truth*, London: George Allen and Unwin.
Niiniluoto, Ilkka (1977), "On A K-Dimensional System of Inductive Logic," in P. Asquith, ed., *PSA 1976*, East Lansing: Philosophy of Science Association, pp. 425-47.
Popper, Karl R. (1959), *The Logic of Scientific Discovery*, London, Hutchinson and Co.
Ramsey, Frank P. ([1931] 1950), "Probability and Partial Belief," in R. B. Braithwaite, ed., *The Foundations of Mathematics and Other Logical Essays by Frank P. Ramsey*, London: Routledge and Kegan Paul, pp. 256-57.
Rescher, Nicholas (1958), "Theory of Evidence," *Philosophy of Science*, 25, pp. 83-94.
Savage, Leonard J. (1954), *The Foundations of Statistics*, New York: John Wiley and Sons.
Shafer, Glenn (1976), *A Mathematical Theory of Evidence*, Princeton: University of Princeton Press.
Shortliffe, E. H. (1976), *Computer-Based Medical Consultations: MYCIN*, New York: Elsevier.
Teller, Paul (1976), "Conditionalization, Observation, and Change of Preference," in W. Harper and C. Hooker, eds., *Foundations of Probability Theory*, Vol. I, Dordrecht, Netherlands: Reidel, pp. 205-53.
Van Fraassen, Bas (1984), "Belief and Will," *Journal of Philosophy*, 81, pp. 235-36.
Venn, John (1866), *The Logic of Chance*, London: Macmillan; reprinted New York: Chelsea, 1963.
Zadeh, Lotfi (1979), "On the Validity of Dempster's Rule of Combination of Evidence," Berkeley, Memo UCB/ERL M79,24.